22

Metaphor in science

THOMAS S. KUHN

If I had been preparing the main paper on the role of metaphor in science, my point of departure would have been precisely the works chosen by Boyd: Max Black's well-known paper on metaphor (Black, 1962b), together with recent essays by Kripke and Putnam on the causal theory of reference (Kripke, 1972; Putnam, 1975a, 1975b). My reasons for those choices would, furthermore, have been very nearly the same as his, for we share numerous concerns and convictions. But, as I moved away from the starting point that body of literature provides, I would quite early have turned in a direction different from Boyd's, following a path that would have brought me quickly to a central metaphorlike process in science, one which he passes by. That path I shall have to sketch, if sense is to be made of my reactions to Boyd's proposals, and my remarks will therefore take the form of an excessively condensed epitome of parts of a position of my own, comments on Boyd's paper emerging along the way. That format seems all the more essential inasmuch as detailed analysis of individual points presented by Boyd is not likely to make sense to an audience largely ignorant of the causal theory of reference.

Boyd begins by accepting Black's "interaction" view of metaphor. However metaphor functions, it neither presupposes nor supplies a list of the respects in which the subjects juxtaposed by metaphor are similar. On the contrary, as both Black and Boyd suggest, it is sometimes (perhaps always) revealing to view metaphor as creating or calling forth the similarities upon which its function depends. With that position I very much agree and, lacking time, I shall supply no arguments for it. In addition, and presently more significant, I agree entirely with Boyd's assertion that the open-

endedness or inexplicitness of metaphor has an important (and I think precise) parallel in the process by which scientific terms are introduced and thereafter deployed. However scientists apply terms like "mass," "electricity," "heat," "mixture," or "compound" to nature, it is not ordinarily by acquiring a list of criteria necessary and sufficient to determine the referents of the corresponding terms.

With respect to reference, however, I would go one step further than Boyd. In his chapter, the claims for a parallel to metaphor are usually restricted to the theoretical terms of science. I suppose that they often hold equally for what used to be called observation terms, for example "distance," "time," "sulphur," "bird," or "fish." The fact that the last of these terms figures large in Boyd's examples suggests that he is unlikely to disagree. He knows as well as I that recent developments in philosophy of science have deprived the theoretical/observational distinction of anything resembling its traditional cash value. Perhaps it can be preserved as a distinction between antecedently available terms and new ones introduced at particular times in response to new scientific discoveries or inventions. But, if so, the parallel to metaphor will hold for both. Boyd makes less than he might of the ambiguity of the word "introduced." Something with the properties of metaphor is often called upon when a new term is *introduced into* the vocabulary of science. But it is also called upon when such terms – by now established in the common parlance of the profession – are *introduced to* a new scientific generation by a generation that has already learned their use. Just as reference must be established for each new element in the vocabulary of science, so accepted patterns of reference must be reestablished for each new cohort of recruits to the sciences. The techniques involved in both modes of introduction are much the same, and they therefore apply on both sides of the divide between what used to be called "observational" and "theoretical" terms.

To establish and explore the parallels between metaphor and reference fixing, Boyd resorts both to the Wittgensteinian notion of natural families or kinds and to the causal theory of reference. I would do the same, but in a significantly different way. It is at this point that our paths begin to diverge. To see how they do so, look first at the causal theory of reference itself. As Boyd notes, that theory originated and still functions best in application to proper names like "Sir Walter Scott." Traditional empiricism suggested that proper names refer by virtue of an associated definite description chosen to provide a sort of definition of the name: for example, "Scott is the author of *Waverley.*" Difficulties immediately arose, because the choice of the defining description seemed arbitrary. Why should being the author of the novel *Waverley* be a criterion governing the applicability of the name "Walter Scott" rather than a historical fact about the individual to whom the name, by whatever techniques, does refer? Why should having written *Waverley* be a necessary characteristic of Sir Walter Scott but having writ-

ten *Ivanhoe* a contingent one? Attempts to remove these difficulties by using more elaborate definite descriptions, or by restricting the characteristics on which definite descriptions may call, have uniformly failed. The causal theory of reference cuts the Gordian knot by denying that proper names have definitions or are associated with definite descriptions at all.

Instead, a name like "Walter Scott" is a tag or label. That it attaches to one individual rather than to another or to no one at all is a product of history. At some particular point in time a particular infant was baptized or dubbed with the name "Walter Scott," which he bore thereafter through whatever events he happened to experience or bring about (for example, writing *Waverley*). To find the referent of a name like "Sir Walter Scott" or "Professor Max Black," we ask someone who knows the individual about whom we inquire to point him out to us. Or else we use some contingent fact about him, like his authorship of *Waverley* or of the paper on metaphor, to locate the career line of the individual who happened to write that work. If, for some reason, we doubt that we have correctly identified the person to whom the name applies, we simply trace his life history or lifeline backward in time to see whether it includes the appropriate act of baptism or dubbing.

Like Boyd, I take this analysis of reference to be a great advance, and I also share the intuition of its authors that a similar analysis should apply to the naming of natural kinds; Wittgenstein's games, birds (or sparrows), metals (or copper), heat, and electricity. There is something right about Putnam's claim that the referent of "electric charge" is fixed by pointing to the needle of a galvanometer and saying that "electric charge" is the name of the physical magnitude responsible for its deflection. But, despite the amount that Putnam and Kripke have written on the subject, it is by no means clear just what is right about their intuition. My pointing to an individual, Sir Walter Scott, can tell you how to use the corresponding name correctly. But pointing to a galvanometer needle while supplying the name of the cause of its deflection attaches the name only to the cause of that particular deflection (or perhaps to an unspecified subset of galvanometer deflections). It supplies no information at all about the many other sorts of events to which the name "electric charge" also unambiguously refers. When one makes the transition from proper names to the names of natural kinds, one loses access to the career line or lifeline which, in the case of proper names, enables one to check the correctness of different applications of the same term. The individuals which constitute natural families do have lifelines, but the natural family itself does not.

It is in dealing with difficulties like this one that Boyd makes what I take to be an unfortunate move. To get around them he introduces the notion of "epistemic access," explicitly abandoning in the process all use of "dubbing" or "baptism" and implicitly, so far as I can see, giving up recourse to ostension as well. Using the concept of epistemic access, Boyd has a number

of cogent things to say both about what justifies the use of a particular scientific language and about the relation of a later scientific language to the earlier one from which it has evolved. To some of his points in this area I shall be returning. But despite these virtues, something essential is lost, I think, in the transition from "dubbing" to "epistemic access." However imperfectly developed, "dubbing" was introduced in an attempt to understand how, in the absence of definitions, the referents of individual terms could be established at all. When dubbing is abandoned or shoved aside, the link it provided between language and the world disappears as well. If I understand Boyd's chapter correctly – something I do not take for granted – the problems to which it is directed change abruptly when the notion of epistemic access is introduced. Thereafter, Boyd seems simply to assume that the adherents of a given theory somehow or other know to what their terms refer. How they can do so ceases to concern him. Rather than extending the causal theory of reference, he seems to have given it up.

Let me therefore attempt a different approach. Though ostension is basic in establishing referents both for proper names and for natural kind terms, the two differ not only in complexity but also in nature. In the case of proper names, a single act of ostension suffices to fix reference. Those of you who have seen Richard Boyd once will, if your memories are good, be able to recognize him for some years. But, if I were to exhibit to you the deflected needle of a galvanometer, telling you that the cause of the deflection was called "electric charge," you would need more than good memory to apply the term correctly in a thunderstorm or to the cause of the heating of your electric blanket. Where natural-kind terms are at issue, a number of acts of ostension are required.

For terms like "electric charge," the role of multiple ostensions is difficult to make out, for laws and theories also enter into the establishment of reference. But my point does emerge clearly in the case of terms that are ordinarily applied by direct inspection. Wittgenstein's example, games, will do as well as another. A person who has watched chess, bridge, darts, tennis, and football, and who has also been told that each of them is a game, will have no trouble in recognizing that both backgammon and soccer are games as well. To establish reference in more puzzling cases – prize fights or fencing matches, for example – exposure is required also to members of neighboring families. Wars and gang rumbles, for example, share prominent characteristics with many games (in particular, they have sides and, potentially, a winner), but the term "game" does not apply to them. Elsewhere I have suggested that exposure to swans and geese plays an essential role in learning to recognize ducks (Kuhn, 1974). Galvanometer needles may be deflected by gravity or a bar magnet as well as by electric charge. In all these areas, establishing the referent of a natural-kind term requires exposure not only to varied members of that kind but also to members of others – to individuals, that is, to which the term might

otherwise have been mistakenly applied. Only through a multiplicity of such exposures can the student acquire what other authors in this book (for example, Cohen and Ortony) refer to as the *feature space* and the knowledge of *salience* required to link language to the world.

If that much seems plausible (I cannot, in a presentation so brief, hope to make it more), then the parallel to metaphor at which I have been aiming may be apparent as well. Exposed to tennis and football as paradigms for the term "game," the language learner is invited to examine the two (and soon, others as well) in an effort to discover the characteristics with respect to which they are alike, the features that render them similar, and which are therefore relevant to the determination of reference. As in the case of Black's interactive metaphors, the juxtaposition of examples calls forth the similarities upon which the function of metaphor or the determination of reference depend. As with metaphor, also, the end product of the interaction between examples is nothing like a definition, a list of characteristics shared by games and only games, or of the features common to both men and wolves and to them alone. No lists of that sort exist (not all games have either sides or a winner), but no loss of functional precision results. Both natural-kind terms and metaphors do just what they should without satisfying the criteria that a traditional empiricist would have required to declare them meaningful.

My talk of natural-kind terms has not yet, of course, quite brought me to metaphor. Juxtaposing a tennis match with a chess game may be part of what is required to establish the referents of "game," but the two are not, in any usual sense, metaphorically related. More to the point, until the referents of "game" and of other terms which might be juxtaposed with it in metaphor have been established, metaphor itself cannot begin. The person who has not yet learned to apply the terms "game" and "war" correctly can only be misled by the metaphor, "War is a game," or "Professional football is war." Nevertheless, I take metaphor to be essentially a higher-level version of the process by which ostension enters into the establishment of reference for natural-kind terms. The actual juxtaposition of a series of exemplary games highlights features which permit the term "game" to be applied to nature. The metaphorical juxtaposition of the terms "game" and "war" highlights other features, ones whose salience had to be reached in order that actual games and wars could constitute separate natural families. If Boyd is right that nature has "joints" which natural-kind terms aim to locate, then metaphor reminds us that another language might have located different joints, cut up the world in another way.

Those last two sentences raise problems about the very notion of joints in nature, and I shall return to them briefly in my concluding remarks about Boyd's view of theory change. But one last point needs first to be made about metaphor in science. Because I take it to be both less obvious and more fundamental than metaphor, I have so far emphasized the metaphor-

like process which plays an important role in finding the referents of scientific terms. But, as Boyd quite rightly insists, genuine metaphors (or, more properly analogies) are also fundamental to science, providing on occasions "an irreplacable part of the linguistic machinery of a scientific theory," playing a role that is "*constitutive* of the theories they express, rather than merely exegetical." Those words are Boyd's, and the examples which accompany them are good ones. I particularly admire his discussion of the role of the metaphors which relate cognitive psychology to computer science, information theory, and related disciplines. In this area, I can add nothing useful to what he has said.

Before changing the subject, however, I would suggest that what Boyd does say about these "constitutive" metaphors may well have a bearing wider than he sees. He discusses not only "constitutive" but also what he calls "exegetical or pedagogical" metaphors, for example those which describe atoms as "miniature solar systems." These, he suggests, are useful in teaching or explaining theories, but their use is only heuristic, for they can be replaced by nonmetaphorical techniques. "One can say," he points out, "*exactly* in what respects Bohr thought atoms were like solar systems without employing any metaphorical devices, and this was true when Bohr's theory was proposed."

Once again, I agree with Boyd but would nevertheless draw attention to the way in which metaphors like that relating atoms and solar systems are replaced. Bohr and his contemporaries supplied a model in which electrons and nucleus were represented by tiny bits of charged matter interacting under the laws of mechanics and electromagnetic theory. That model replaced the solar system metaphor but not, by doing so, a metaphorlike process. Bohr's atom model was intended to be taken only more-or-less literally; electrons and nuclei were not thought to be exactly like small billiard or Ping-Pong balls; only some of the laws of mechanics and electromagnetic theory were thought to apply to them; finding out which ones did apply and where the similarities to billiard balls lay was a central task in the development of the quantum theory. Furthermore, even when that process of exploring potential similarities had gone as far as it could (it has never been completed), the model remained essential to the theory. Without its aid, one cannot even today write down the Schrödinger equation for a complex atom or molecule, for it is to the model, not directly to nature, that the various terms in that equation refer. Though not prepared here and now to argue the point, I would hazard the guess that the same interactive, similarity-creating process which Black has isolated in the functioning of metaphor is vital also to the function of models in science. Models are not, however, merely pedagogic or heuristic. They have been too much neglected in recent philosophy of science.

I come now to the large part of Boyd's chapter that deals with theory choice, and I shall have to devote disproportionately little time to my

discussion of it. That may, however, be less of a drawback than it seems, for attention to theory choice will add nothing to our central topic, metaphor. In any case, with respect to the problem of theory change, there is a great deal about which Boyd and I agree. And in the remaining area, where we clearly differ, I have great difficulty articulating just what we disagree about. Both of us are unregenerate realists. Our differences have to do with the commitments that adherence to a realist's position implies. But neither of us has yet developed an account of those commitments. Boyd's are embodied in metaphors which seem to me misleading. When it comes to replacing them, however, I simply waffle. Under these circumstances, I shall attempt only a rough sketch of the areas in which our views coincide and in which they appear to diverge. For the sake of brevity in that attempt, furthermore, I shall henceforth drop the distinction on which I have previously insisted between metaphor itself and metaphorlike processes. In these concluding remarks, "metaphor" refers to all those processes in which the juxtaposition either of terms or of concrete examples calls forth a network of similarities which help to determine the way in which language attaches to the world.

Presupposing what has already been said, let me summarize those portions of my own position with which I believe Boyd largely agrees. Metaphor plays an essential role in establishing links between scientific language and the world. Those links are not, however, given once and for all. Theory change, in particular, is accompanied by a change in some of the relevant metaphors and in the corresponding parts of the network of similarities through which terms attach to nature. The earth was like Mars (and was thus a planet) after Copernicus, but the two were in different natural families before. Salt-in-water belonged to the family of chemical compounds before Dalton, to that of physical mixtures afterwards. And so on. I believe, too, though Boyd may not, that changes like these in the similarity network sometimes occur also in response to new discoveries, without any change in what would ordinarily be referred to as a scientific theory. Finally, these alterations in the way scientific terms attach to nature are not – logical empiricism to the contrary – purely formal or purely linguistic. On the contrary, they come about in response to pressures generated by observation or experiment, and they result in more effective ways of dealing with some aspects of some natural phenomena. They are thus substantive or cognitive.

These aspects of Boyd's and my agreement should occasion no surprise. Another one may, though it ought not. Boyd repeatedly emphasizes that the causal theory of reference or the concept of epistemic access makes it possible to compare successive scientific theories with each other. The opposing view, that scientific theories are incomparable, has repeatedly been attributed to me, and Boyd himself may believe I hold it. But the book on which this interpretation is imposed includes many explicit exam-

ples of comparisons between successive theories. I have never doubted either that they were possible or that they were essential at times of theory choice. Instead, I have tried to make two rather different points. First, comparisons of successive theories with each other and with the world are never sufficient to dictate theory choice. During the period when actual choices are made, two people fully committed to the values and methods of science, and sharing also what both concede to be data, may nevertheless legitimately differ in their choice of theory. Second, successive theories are incommensurable (which is not the same as incomparable) in the sense that the referents of some of the terms which occur in both are a function of the theory within which those terms appear. There is no neutral language into which both of the theories as well as the relevant data may be translated for purposes of comparison.

With all of this I believe, perhaps mistakenly, that Boyd agrees. If so, then our agreement extends one step further still. Both of us see in the causal theory of reference a significant technique for tracing the continuities between successive theories and, simultaneously, for revealing the nature of the differences between them. Let me provide an excessively cryptic and simplistic example of what I, at least, have in mind. The techniques of *dubbing* and of *tracing lifelines* permit astronomical individuals – say, the earth and moon, Mars and Venus – to be traced through episodes of theory change, in this case the one due to Copernicus. The lifelines of these four individuals were continuous during the passage from heliocentric to geocentric theory, but the four were differently distributed among natural families as a result of that change. The moon belonged to the family of planets before Copernicus, not afterwards; the earth to the family of planets afterwards, but not before. Eliminating the moon and adding the earth to the list of individuals that could be juxtaposed as paradigms for the term "planet" changed the list of features salient to determining the referents of that term. Removing the moon to a contrasting family increased the effect. That sort of redistribution of individuals among natural families or kinds, with its consequent alteration of the features salient to reference, is, I now feel, a central (perhaps the central) feature of the episodes I have previously labeled scientific revolutions.

Finally, I shall turn very briefly to the area in which Boyd's metaphors suggest that our paths diverge. One of those metaphors, reiterated throughout his chapter, is that scientific terms "cut [or can cut] nature at its joints." That metaphor and Field's notion of quasi-reference figure large in Boyd's discussion of the development of scientific terminology over time. Older languages succeeded, he believes, in cutting the world at, or close to, some of its joints. But they also often committed what he calls "real errors in classification of natural phenomena," many of which have since been corrected by "more sophisticated accounts of those joints." The older language may, for example, "have classified together certain things which

have no important similarity, or [may] have failed to classify together, things which are, in fact, *fundamentally* similar" (italics added). This way of talking is, however, only a rephrased version of the classical empiricists' position that successive scientific theories provide successively closer approximations to nature. Boyd's whole chapter presupposes that nature has one and only one set of joints to which the evolving terminology of science comes closer and closer with time. At least, I can see no other way to make sense of what he says in the absence of some theory-independent way of distinguishing *fundamental* or *important* similarities from those that are *superficial* or *unimportant*.[1]

To describe the successive-approximation view of theory change as a presupposition does not, of course, make it wrong, but it does point to the need for arguments missing from Boyd's paper. One form such arguments might take is the empirical examination of a succession of scientific theories. No pair of theories will do, for the more recent could, by definition, be declared the better approximation. But, given a succession of three or more theories directed to more-or-less the same aspects of nature, it should be possible, if Boyd is right, to display some process of bracketing and zeroing in on nature's real joints. The arguments which would be required are both complex and subtle. I am content to leave open the question to which they are directed. But my strong impression is that they will not succeed. Conceived as a set of instruments for solving technical puzzles in selected areas, science clearly gains in precision and scope with the passage of time. As an instrument, science undoubtedly does progress. But Boyd's claims are not about the instrumental effectiveness of science but rather about its ontology, about what really exists in nature, about the world's real joints. And in this area I see no historical evidence for a process of zeroing in. As I have suggested elsewhere, the ontology of relativistic physics is, in significant respects, more like that of Aristotelian than that of Newtonian physics. That example must here stand for many.

Boyd's metaphor of nature's joints relates closely to another, the last I shall attempt to discuss. Again and again, he speaks of the process of theory change as one which involves "the accommodation of language to the world." As before, the thrust of his metaphor is ontological; the world to which Boyd refers is the one real world, still unknown but toward which science proceeds by successive approximation. Reasons for being uneasy with that point of view have already been described, but this way of expressing the viewpoint enables me to phrase my reservations in a different way. What is the world, I ask, if it does not include most of the sorts of things to which the *actual* language spoken at a given time refers? Was the earth really a planet in the world of pre-Copernican astronomers who spoke a language in which the features salient to the referent of the term "planet" excluded its attachment to the earth? Does it obviously make better sense to speak of accommodating language to the world than of accommodating

THOMAS S. KUHN

the world to language? Or is the way of talking which creates that distinction itself illusory? Is what we refer to as "the world" perhaps a product of a mutual accommodation between experience and language?

I shall close with a metaphor of my own. Boyd's world with its joints seems to me, like Kant's "things in themselves," in principle unknowable. The view toward which I grope would also be Kantian but without "things in themselves" and with categories of the mind which could change with time as the accommodation of language and experience proceeded. A view of that sort need not, I think, make the world less real.

NOTE

1 In revising the manuscript to which this paragraph and those following are addressed, Boyd has pointed out that both natural kinds and nature's joints may be context- or discipline- or interest-relative. But, as note 2 to his paper will indicate, that concession does not presently bring our positions closer together. It may do so in the future, however, for the same root note undermines the position it defends. Boyd concedes (mistakenly, I think) that a kind is "un-'objective' " to the extent that it is context- or discipline-dependent. But that construal of "objective" requires that context-independent bounds be specified for context-dependence. If any two objects could, in principle, be rendered similar by choice of an appropriate context, then objectivity, in Boyd's sense, would not exist. The problem is the same as the one suggested by the sentence to which this footnote is attached.